**SCENARIO-BASED QUESTIONS**

1. You are working on an NLP model. So, you are dealing with words and sentences, not numbers. Your problem is to categorize these words and make sense of them. Your manager told you that you have to use embeddings. Which of the following techniques are not related to embeddings? Explain the other terms related to embedding.

A. Count Vector

B. TF-IDF Vector

C. Co-Variance Matrix

1. You are a junior Data Scientist and are working on a deep neural network model to optimize the level of customer satisfaction for after-sales services with the goal of creating greater client loyalty. You are struggling with your model (learning rates, hidden layers and nodes selection) for optimizing processing and to let it converge in the fastest way. What is this problem called in ML language?
2. Imagine, you are working with a microblogging website and you want to develop a machine learning algorithm which predicts the number of views on the articles. Your analysis is based on features like author name, number of articles written by the same author in past and a few other features. Which evaluation metric would you choose in that case? Explain the metric used.
3. You are working with categorical feature(s) and you have not looked at the distribution of the categorical variable in the test data. You want to apply one hot encoding (OHE) on the categorical feature(s). What challenges you may face if you have applied OHE on a categorical variable of train dataset?
4. You’ve built a random forest model with 10000 trees. You got delighted after getting training error as 0.00. But, the validation error is 34.23. What is going on? Haven’t you trained your model perfectly?
5. We know that one hot encoding increases the dimensionality of a data set. But, label encoding doesn’t. How?
6. A company is looking to improve their fraud detection system. How can they use an artificial neural network to identify potentially fraudulent transactions?
7. A researcher wants to develop a system that can predict whether a cancer patient is likely to respond well to a particular treatment. How can they train an artificial neural network to make these predictions?
8. An automotive company is developing a self-driving car. How can they use an artificial neural network to recognize traffic signs and signals?
9. A bank wants to develop a chatbot that can assist customers with their banking needs. How can they use an artificial neural network to understand and respond to customer inquiries?
10. A company is trying to reduce energy consumption in their buildings. How can they use an artificial neural network to optimize temperature control and lighting schedules?
11. A retailer wants to develop a recommendation engine that suggests products to customers based on their purchase history. How can they use an artificial neural network to make these recommendations?
12. A university is developing a system that can automatically grade student essays. How can they use an artificial neural network to evaluate the quality of the essays and assign grades?
13. A bank is using a classification algorithm to identify fraudulent transactions. However, the algorithm is incorrectly flagging some legitimate transactions as fraudulent. How can the bank improve the efficiency of the classification algorithm while still maintaining a high level of accuracy?
14. A healthcare provider is using a classification model to predict whether patients are at risk of developing a particular condition. However, the model has a high false negative rate, meaning that many at-risk patients are not being identified. What steps can the provider take to improve the sensitivity of the model without sacrificing specificity?
15. A marketing team is using a classification algorithm to identify potential customers who are most likely to purchase their product. However, the algorithm is producing a large number of false positives, meaning that many customers who are not likely to purchase are being identified as high-potential leads. How can the team improve the precision of the algorithm without sacrificing recall?
16. A transportation company is using a classification model to predict whether a shipment is likely to be delayed. However, the model has a high false positive rate, meaning that many shipments that are not actually at risk of delay are being flagged as such. How can the company improve the efficiency of the model while still maintaining a high level of accuracy?
17. An e-commerce website is using a classification algorithm to recommend products to customers based on their browsing and purchase history. However, the algorithm is not taking into account the seasonal fluctuations in customer preferences, resulting in less relevant recommendations during certain times of the year. How can the website improve the efficiency of the algorithm by incorporating seasonal trends into its predictions?
18. A real estate agent wants to predict the selling price of a house based on its features, such as size, location, and number of bedrooms. How can a regression model be used to predict the selling price of the house, and what features should be included in the model?
19. A manufacturer wants to predict the amount of time it will take to produce a particular product based on its complexity, the number of steps involved in the production process, and the experience level of the workers. How can a regression model be used to predict production time, and what factors should be included in the model?
20. A financial institution wants to predict the credit risk of a loan applicant based on their income, credit score, and employment history. How can a regression model be used to predict credit risk, and what factors should be included in the model?
21. A transportation company wants to predict the fuel efficiency of its vehicles based on factors such as the weight of the cargo, the distance traveled, and the terrain. How can a regression model be used to predict fuel efficiency, and what factors should be included in the model?
22. A university wants to predict the GPA of its students based on factors such as their SAT scores, high school GPA, and extracurricular activities. How can a regression model be used to predict GPA, and what factors should be included in the model?
23. A marketing team wants to group their customers based on their buying habits in order to develop targeted marketing campaigns. How can clustering be used to segment customers, and what features should be used to cluster them?
24. A manufacturer wants to group their suppliers based on their delivery times and product quality in order to identify the most reliable suppliers. How can clustering be used to group suppliers, and what features should be used to cluster them?
25. A healthcare provider wants to group their patients based on their medical history and symptoms in order to identify common conditions and develop more effective treatments. How can clustering be used to group patients, and what features should be used to cluster them?
26. A social media platform wants to group their users based on their interests and behavior in order to provide more relevant content and advertisements. How can clustering be used to group users, and what features should be used to cluster them?
27. A transportation company wants to group their delivery routes based on their traffic patterns and delivery times in order to optimize their delivery schedules. How can clustering be used to group routes, and what features should be used to cluster them?
28. You work as a machine learning specialist for a consulting firm where you analyze data about the consultants who work there in preparation for using the data in your machine learning models. The features you have in your data are things like employee id, specialty, practice, job description, billing hours, and principle. The principle attribute is represented as ‘yes’ or ‘no’, whether the consultant has made principle level or not. For your initial analysis, you need to identify the distribution of consultants and their billing hours for the given period. What visualization best describes this relationship?
29. Consider a manufacturing firm that has a dataset containg the expenditure made on advertisements in different media such as TV, Radio and Newspaper along with the Amount of sales made. The firm wants to optimize the Amount of sales based on the above factors. You are asked to :

a) create a model that quantitatively relates the sales with the amount of expenditure on advertisement via TV, radio and newspaper.

b) To know the accuracy of the model i.e., how well these variables can predict the amount of sales

Discuss the steps followed to build such a model. Recommend the suitable techniques to consider at each step.

1. Your work for a company that performs seismic research for client firms that drill for petroleum. As a machine learning specialist, you have built a series of models that classify seismic waves to determine the seismic profile of a proposed drilling site. You need to select the best model to use in production. Which metric should you use to compare and evaluate your machine learning classification models against each other?
2. You work for the security department of your firm. As part of securing your firm’s email activity from phishing attacks, you need to build a machine learning model that analyzes incoming email text to find word phrases like “you’re a winner” or “click here now” to find potential phishing emails. Which text feature engineering techniques is the best solution for this task? Justify